Validation of Clinical vs Algorithmic Definition of Line of Therapy in Multiple Myeloma

James Black[1]; Uzor Ogbu[1]; Elisabeth Wassner Fritsch[2]; John Byon[3]; Jingbo Yi[4]; Mike Lu[3]

[1]PHC Data Science, Hoffman-La Roche, Basel, Switzerland [2]Clinical Science, Hoffman-La Roche, Basel, Switzerland

[3]Clinical Science, Genentech, South San Francisco, United States; John Byon left Genentech during this research project [4]Genesis Research, New Jersey, United States

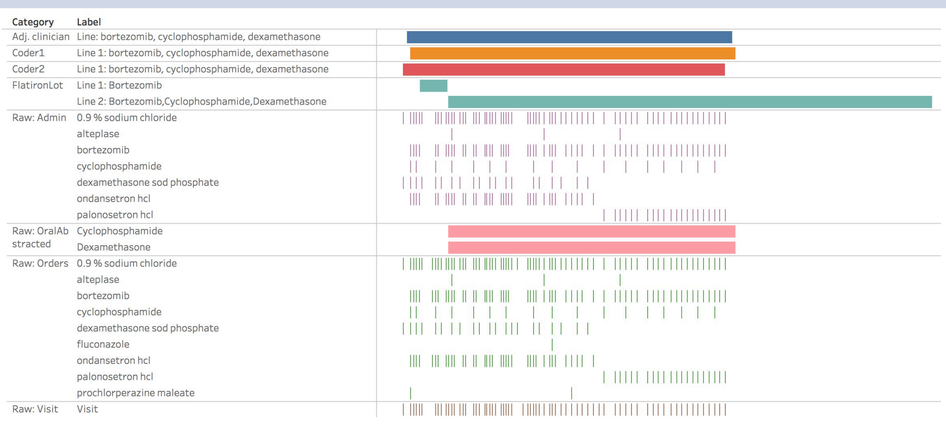
Background

- Electronic health records (EHR) enable the passive collection of real world data. Algorithms exist to convert recorded treatment events into clinically relevant line of therapies (LoTs) without the need for human abstraction and coding. While the EHR provides a rich profile of an oncology patient's treatment history, it is important to validate the routinely used algorithmic LoT defined by data providers.

Methods

- 100 Multiple Myeloma (MM) patients initiating first line (1L) on or after 2011-01-01 where sampled from the Flatiron Health EHR-derived database[1], a a longitudinal, demographically and geographically diverse.
- Algorithm based LoT was developed by Flatiron Health and uses raw administrations, medication orders and abstracted information on specific oral medications relevant to the indication.
- Two clinicians independently defined LoTs for each patient using EHR data combined with information on abstracted transplants and oral therapies.
- Clinicians defined LoT according to their interpretation of NCCN[2] guidelines, and relevant IMWG[3] statements.
- A third clinician reconciled any discrepancies between clinicians before LoT definitions were compared between clinicians and the Flatiron LoT algorithm.
- In addition to subjective assessment of the concordance, Fleiss' kappa was used to assess observed vs expected agreement in LoTs and Bland-Altman plots, histograms and Intraclass coefficients to assess differences in duration where LoTs concur.

Figure 1: Example patient with raw data and different LoT estimates

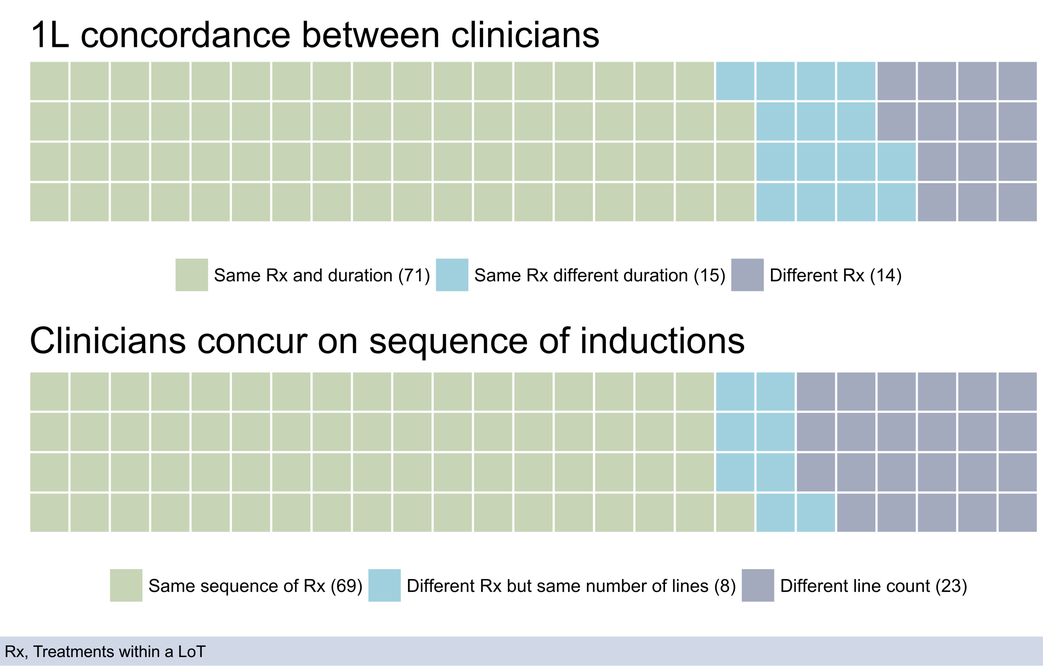


Adj., adjucated; FlatironLoT, Algorithm LoT; Orders, Medication was ordered for use on this date; Admin, medication was administered on this date

Results – clinician vs clinician

- Clinicians identified the same 1L induction in 86 (86%) of patients; Kappa statistic 0.83 (95%CI 0.75,0.91). Within the 46 patients where at least one clinician said there was a 2L, they reported the same 2L for 26 (57%) of the patients; Kappa statistic 0.63 (95%CI 0.47,0.78).
- In the 86 with the same 1L, the duration of the 1L line was identical for 71 (83%) patients. For the 86 subjects, rated by 2 raters, the intraclass correlation coefficient is 0.959 (95%CI 0.938,0.973).
- Agreement on same induction sequence across all lines (i.e. 1L, 2L, …) was achieved in 69 (69%) of patients, while total count of lines a patient experienced was the same in 77 (77%) of patients.

Figure 2: Comparing concordance between clinicians (n=100)

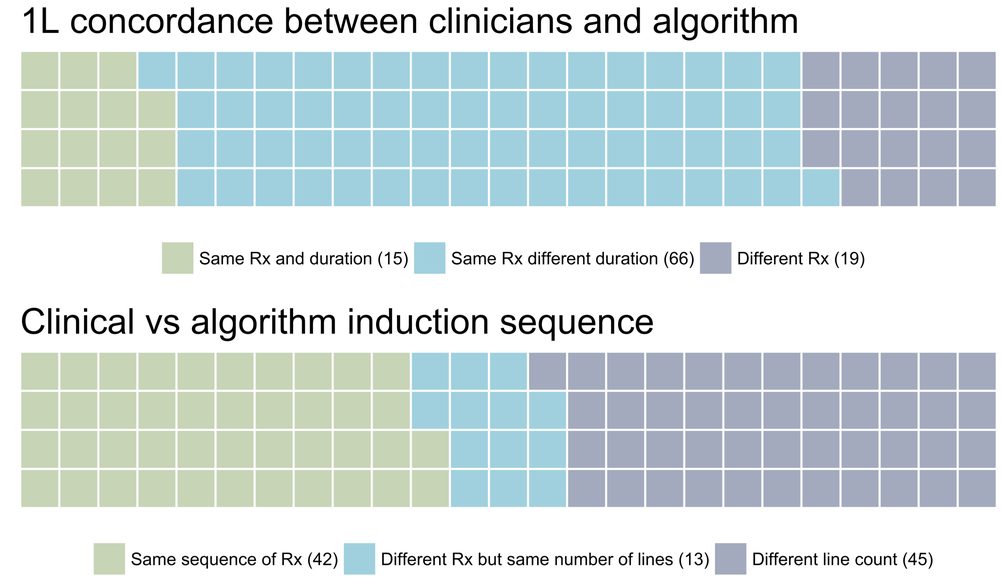


Rx, Treatments within a LoT

Results – clinicians vs Flatiron algorithm

- Where clinicians disagreed, a third clinician adjudicated the recorded lines. Comparing the reconciled clinical lines against the algorithmic lines, 81 (81%) of patients had the same first line induction; Kappa statistic 0.77 (95%CI 0.67,0.86). Within the 44 patients where reconciled clinician LoTs or the algorithm identified a 2L, they reported the same 2L for 19 (43%) patients; Kappa statistic 0.49 (95%CI 0.32,0.66).
- In the 81 with the same 1L, the duration of the 1L line was identical for 15 (19%) patients. For the 81 subjects, rated by 2 raters, the intraclass correlation coefficient is 0.52 (95%CI 0.34,0.66).
- Agreement on same induction sequence across all lines achieved in 42 (42%) patients, while total count of lines a patient experienced was the same in 55 (55%) patients.

Figure 3: Clinicians vs algorithm concordance (n = 100)



Rx, Treatments in LoT

Results – clinicians vs Flatiron algorithm

Table 1: Reason* for dis-concordance between clinicians and algorithm in 1L

	Different 1L treatment (n = 19)
Drug was an non-admin order, and not used by clinicians	9
Drug added >30 days of initiation so algorithm did not capture as part of same line, while clinicians did	5
Algorithm did not consider an oral drug order	3
Clinicians noted a change in line during first 28 days	1
Clinicians did not consider a drug	1

Where components differed. *Subjective assessment made by agreement between researchers (Black and Yi)

Conclusions

- Defining LoT based on observed EHR profiles introduces a degree of uncertainty into classifications. When seen in the context of inter-rater agreement between clinicians, algorithmic performance is comparable to clinician defined lines, while allowing the processing of LoT classifications at scale.
- LoT definitions become less accurate in 2L, relative to 1L, suggesting LoT is less reliable in relapse populations.
- Guidelines and working group statements did not provide a definitive guidance to define periods of induction, maintenance and consolidation in MM. Nonetheless, these generally held for the breadth of treatment experiences seen in the real world.
- Rule-based LoT methods differed most when clinicians applied clinical judgement to decide whether an order with no evidence of administration was cancelled, despite the order not being cancelled in the EHR.
- LoT analyses would benefit from the collection of clinical intent either from medical record abstraction or collection within the EHR system.

Conflicts of interest

All authors are employees of, or contractors to, the Roche Group. Flatiron Health is a subsidiary of the Roche Group.

References

1. Flatiron Health database (https://flatiron.com/real-world-evidence/), September 2017
2. NCCN Clinical Practice Guidelines in Oncology. (2016) https://doi.org/10.1007/978-3-540-85772-3
3. IMWG consensus on maintenance therapy in multiple myeloma. (2012) https://doi.org/10.1182/blood-2011-11-374249.

Acknowledgements

Dr Adrian Cassidy for reviewing drafts.

Roche

Genentech

A Member of the Roche Group

34th International Conference on Pharmacoepidemiology and Therapeutic Risk Management
Prague, Czech Republic | August 22 – August 26, 2018